Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v8n3p384

Some considerations to carry out a composite indicator for ordinal data

By Zanarotti, Pagani

Published: 20 December 2015

Electronic Journal of Applied Statistical Analysis
Vol. 08, Issue 03, December 2015, 384-397
DOI: 10.1285/i20705948v8n3p384

Some considerations to carry out a composite indicator for ordinal data

Maria Chiara Zanarotti*[a] and Laura Pagani[b]

[a] *Catholic University of Milan, Department of Statistical Sciences, Largo Gemelli, 1, 20123 Milan, Italy*
[b] *University of Udine, Department of Economics and Statistics, Via Tomadini, 30/a, 33100 Udine, Italy*

Published: 20 December 2015

Composite indicators (CIs) are important and useful tools in many fields to assess, compare and rank performances, development stage, quality and many other different targets. CIs are an overall measure of a multidimensional, not directly observable, concept and are obtained by means of a set of manifest variables (elementary indicators) that contribute to define the overall measure. In this paper, some matters regarding methods to build CIs are reviewed, assuming elementary indicators are ordinal and quantification is necessary to convert observed data into a numerical form. Scoring methods, aggregating functions and weighting systems are considered. In particular, a scoring method based on the observed distribution or the use of dissimilarity indices for quantification together with the Kendall-τ association or a heterogeneity measure for weighting are suggested. Some of the reviewed procedures are compared using students' satisfaction data.

keywords: ordered categorical variables, dissimilarity indices, combining functions, composite indicator, performance index.

1 Introduction

The necessity frequently arises in many fields to obtain, from values of some observable variables, one or some synthetic indices able to summarize all information from manifest variables. Composite indicators (CIs) are the result of this synthesis. Observable variables, called *elementary (sub-) indicators*, can be considered pieces of information that

*Corresponding author: chiara.zanarotti@unicatt.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

summarize the characteristics of a system or highlight what is happening in a system (Saisana and Tarantola, 2002). According to the OECD Glossary of Statistical Terms (2008), "*a CI is formed when individual indicators are compiled into a single index on the basis of an underlying model of the multi-dimensional concept that is being measured*". Technically, a CI is a function of the elementary indicators representing different dimensions of a concept, the description of which is the objective of the study (Saisana *et al.*, 2005). Often, particularly in some fields, like sociology or psychometrics, the multidimensional aspect of interest is considered as a latent variable or latent trait because it is an abstract concept that is not directly observable. It is possible to obtain a measure of the latent variable using a statistical model connecting the latent (unobservable) variable to observed ones (Skrondal and Rabe-Hesketh, 2007). This measure process implies a measure error component. In this paper, we consider the use of observable variables to measure an unobservable aspect in a general framework by the construction of CIs. As already stated, the applicative fields in which the determination of CIs is required are the most varied: from sociology to psychometrics, from economy to development, from policy to *customer/user* satisfaction. Particularly in the latter area, many methodological studies have been recently added (see, for example, Kennet and Salini, 2012) along with the significant request of performance evaluation from private and public offices regarding provided services. In this paper, the focus is on CI construction when indicators are ordinal variables and the applicative context we refer to is that of customer/user satisfaction. Section 2 is devoted to general matters about CI. In Section 3, different ways to transform ordinal categorical data into quantitative data are analyzed, and particular emphasis is given to the use of the cumulative function to obtain such quantification. Section 4 deals with the problem of how to aggregate information derived from many elementary indicators. An application is presented in Section 5, and concluding remarks are reported in Section 6.

2 Methodological aspects

Many steps must be addressed to build a CI. Each of these steps contributes significantly to the "quality" of the final synthetic indicator. These steps can be summarized as follows (Lauro and Nappo, 2011): 1. elementary indicator selection; 2. elementary indicator pretreatment; 3. choice of a system of weights; 4. choice of an aggregation function of elementary indicators; 5. validation. Regarding the first step, the choice of information relevant to take into account the multidimensionality of the unobservable variable depends on the applicative context. The choice can be supported *ex-ante* by appropriate exploratory analysis, and confirmed/modified *ex-post* by validation and calibration procedures. Referring to the wide literature about validation and calibration techniques and about models for latent variables (Bartholomew, 1987; Bollen, 1989; Borsboom *et al.*, 2003), the questions regarding choice of elementary indicators and validation are beyond the scope of this study. To solve steps 2, 3 and 4, it is possible to adopt different approaches that can be broadly classified into two principal types: (a) a model-based approach that makes use of statistical models to solve steps 2–4; (b) a model-free approach that handles in sequence steps 2 up to 4. Also, the type of elementary indicators discriminates between different techniques. If indicators are measured on a quantitative scale, it is possible to consider a large class of statistical models in case (a), while only pretreatment is necessary to make them aggregable in case (b). On the other hand, when indicators are categorical, there are further difficulties related to the fact that many statistical models suppose a cardinal level of an observed variable in case (a), and also every aggregating function in case (b) implies that its arguments are numbers, not categories. So, with such kind of data, only a model for categorical variables are suitable in case (a), or scaling techniques have to be used to obtain numerical values in case (b). Figure 1 summarizes these different cases, named, respectively, 1, 2, 3 and 4.

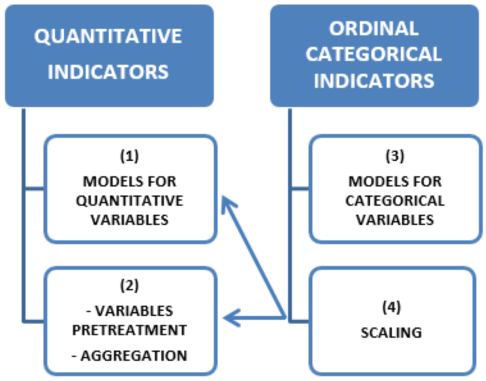

Figure 1: Nature of indicators and aggregation strategy

In this study our attention is devoted only to the case in which elementary indicators are ordinal variables, so only to cases (3) and (4). In case (3) it is possible to use models devoted to qualitative variables, such as latent variable models with ordinal variables (Gifi, 1990; Joreskog and Moustaki, 2001), Rasch model and the family of Item Response Theory (Rasch, 1960), or Combination of Uniform and (shifted) Binomial random variables (CUB) models (D'Elia and Piccolo, 2005; Piccolo, 2009; Iannario and Piccolo, 2010). Otherwise, the alternative is to build the CI after a scaling (case 4) of each ordinal indicator. Scaling can be realized with different procedures, some of which are listed in the next section. After the scaling procedures, a CI is obtained, as in case (2), through an aggregating function where all scaled indicators are weighted with a set of weights.

3 From ordinal to quantitative data

Suppose n judges give a response to J items and that for each item there are K ordered responses, *i.e.* evaluations are collected by a Likert-type scale. Each item is supposed to measure one aspect of the unobservable variable, and the goal is to aggregate all these responses (elementary indicators) to obtain a unique value useful to make a comparison or to establish a ranking. To allow aggregation of responses, it is necessary to handle each item with a numerical indicator. This can be performed in two different ways:

a. scoring ordinal responses and then using these quantitative scores to obtain the CI;

b. obtaining an index suitable for further aggregation from the observed response distribution of each item.

In case a, there are many methods that have been proposed to quantify ordinal variables. Of particular note are the following: (1) linear assignment of the first positive integers to the original ordinal categories; (2) indirect quantification based on theoretical cumulative function (Thurstone, 1928; Torgerson,1958; Portoso, 2003); (3) dual scaling (Nishisato, 1980); (4) homogeneity analysis of the Gifi System (Gifi, 1990); (5) fuzzy quantification method (Zani *et al.*, 2012); (6) quantification using the observed cumulative function (Zani and Berzieri, 2008); (7) score transformation using linear assignment plus observed relative frequency distribution (Arboretti Giancristofaro *et al.*, 2007). In case b, scoring of the original ordinal categories is not necessary, and the solution is to select a suitable index able to measure, for each item, the evaluations of all judges. One index that meets this need is a dissimilarity index, as suggested by Capursi and Porcu (2001). The next subsections are devoted to this subject. In particular for case a, we choose to consider method (7) and, for case b, dissimilarity indices.

3.1 Score transformation and extreme profiles

Without loss of generality, we suppose that items describe different aspects of satisfaction of consumers or users and that each item has the same even number of response categories.

Let us consider the observed ordinal response variable $X_j,\ j = 1, 2, \ldots, J$ with K categories or levels of satisfaction, related to item (aspect) j. We denote with f_{jk} the relative frequency of category k ($k = 1, 2, \ldots, K$) for item j and with F_{jk} the corresponding relative cumulative frequency.

Following Arboretti Giancristofaro *et al.* (2007), for each item j we assign values $v_1, v_2, \ldots, v_k, \ldots, v_K$ (for example $v_k = k$) to the K levels of satisfaction and to separate the low levels of satisfaction or dissatisfaction judgements, *i.e.* $1, 2, \ldots, k_t$ from the high levels of satisfaction or satisfaction judgements, *i.e.* $k_t + 1, k_t + 2, \ldots, K$ with K even.

The idea of this approach is to take into account the observed distribution of responses to items favouring high levels of satisfaction and penalizing low levels of satisfaction. In particular, high levels of satisfaction scores are increased proportionally to relative frequencies while low levels of satisfaction are decreased proportionally to relative frequencies.

We describe two kinds of transformations: the "symmetric scoring system" and the "asymmetric scoring system". For item j the "symmetric scoring system" is:

$$\begin{cases} s^S_{jk} = v_k + f_{jk} \times 0.5 & \text{if } k = k_t + 1, k_t + 2, \ldots, K \\ s^S_{jk} = v_k - f_{jk} \times 0.5 & \text{if } k = 1, 2, \ldots, k_t \end{cases} \tag{1}$$

and the elementary item indicator is the weighted average of the new scores:

$$\bar{s}_j^S = \sum_{k=1}^{K} s_{jk}^S f_{jk}, \tag{2}$$

while the "asymmetric scoring system" is:

$$\begin{cases} s_{jk}^A = v_k + f_{jk} \times 0.5 & \text{if } k = k_t + 1, k_t + 2, \ldots, K \\ s_{jk}^A = v_k + (1 - f_{jk}) \times 0.5 & \text{if } k = 1, 2, \ldots, k_t \end{cases} \tag{3}$$

and the elementary item indicator is the weighted average of the new scores:

$$\bar{s}_j^A = \sum_{k=1}^{K} s_{jk}^A f_{jk}. \tag{4}$$

Note that if K is odd the central k'-*th* category usually is the 'neutral' one (*e.g.* neither satisfied nor dissatisfied). In this case, the score is neither favourite nor penalized, *i.e.* $s_{jk'}^S = s_{jk'}^A = v_{k'}$.

Again following Arboretti Giancristofaro *et al.* (2007), we define the "strong performance profile" corresponding to

1. the degenerate distribution of the minimum level of performance with:

$$F_{j1} = F_{j2} = \ldots = F_{j(K-1)} = F_{jK} = 1, \quad j = 1, 2, \ldots, J; \tag{5}$$

2. the degenerate distribution of the maximum level of performance with:

$$F_{j1} = F_{j2} = \ldots = F_{j(K-1)} = 0, F_{jK} = 1 \quad j = 1, 2, \ldots, J. \tag{6}$$

In order compare items, we consider further transformations of item indicators (2) and (4). These functions must be monotonically increasing with respect to the level of satisfaction, that is, tend or equal to 0 for distribution (5) and tend or equal to one for distribution (6).

We consider three different transformation functions:

a. comparison with the minimum (lowest level of performance):

$$\bar{s}_j^{*S} = 1 - \frac{\bar{s}_{min,j}^S}{\bar{s}_j^S}, \quad \bar{s}_j^{*A} = 1 - \frac{\bar{s}_{min,j}^A}{\bar{s}_j^A}; \tag{7}$$

b. comparison with the maximum (highest level of performance):

$$\bar{s}_j^{*S} = \frac{\bar{s}_j^S}{\bar{s}_{max,j}^S}, \quad \bar{s}_j^{*A} = \frac{\bar{s}_j^A}{\bar{s}_{max,j}^A}; \tag{8}$$

c. comparison with the minimum and maximum or normalization (highest and lowest level of performance):

$$\bar{s}_j^{*S} = \frac{\bar{s}_j^S - \bar{s}_{min,j}^S}{\bar{s}_{max,j}^S - \bar{s}_{min,j}^S}, \quad \bar{s}_j^{*A} = \frac{\bar{s}_j^A - \bar{s}_{min,j}^A}{\bar{s}_{max,j}^A - \bar{s}_{min,j}^A}. \tag{9}$$

Note that the values of minimum and maximum depend on the number of satisfaction levels and on the set of values $v_1, v_2, \ldots, v_k, \ldots, v_K$.

3.2 Dissimilarity indices for ordinal variables and reference distribution

Before introducing a dissimilarity index, we define what we mean by similarity between distributions referring to an ordinal categorical variable.

Def. 1. *Given an ordinal categorical variable, two distributions are similar if they have the same categories with associated equal relative cumulative frequencies.*

Since the goal in this study is to measure, for every item j, how n judges quote the aspect related to each item, the observed distribution is compared with a theoretical one chosen as reference to see if they are similar. As a reference distribution, it is possible to choose between different alternatives:
- distribution defined in (5) called minimum performance, as suggested by Zanarotti (2012);
- distribution defined in (6) called maximum performance, as suggested by Capursi and Porcu (2001);
- uniform distribution, called *totally indifferent distribution*, where all relative cumulative frequencies are equal to k/K for every k, as suggested by Zanarotti (2012).

To compare an observed distribution with a reference one, a dissimilarity index can be used (see, for example, Leti, 1983). In particular, in this paper we consider the following index:

$$Z = \frac{1}{K-1} \sum_{k=1}^{K-1} |F_k^o - F_k^r|, \tag{10}$$

taking values in the interval $[0, 1]$, where F_k^o and F_k^r are relative cumulative frequencies of k^{th} category ($k = 1, \ldots, K$), respectively, for the observed distribution and for the reference distribution. Taking (5) as the reference distribution, the index (10) is equal to:

$$IS^1 = \frac{1}{K-1} \sum_{k=1}^{K-1} |F_k - 1|, \tag{11}$$

assuming increasing values as judgements increase, *i.e.* the further the observed distribution is far from the minimum performance one.

Otherwise, if the reference distribution is the maximum performance one given in (6), Capursi and Porcu (2001) suggested taking the ones' complement of the index (10) to obtain a dissimilarity index that increases as the observed distribution approaches the maximum performance one:

$$IS^2 = 1 - \frac{1}{K-1} \sum_{k=1}^{K-1} F_k. \tag{12}$$

It is trivial to verify that index (11) and (12) are equivalent.

Taking as reference distribution the uniform one, Zanarotti (2012) suggested using the following index:

$$IS^3 = \sum_{k=1}^{K-1} \left(\frac{k}{K} - F_k \right) = \frac{K-1}{2} - \sum_{k=1}^{K-1} F_k, \tag{13}$$

obtained by considering the simple sum (up to $K-1$) of differences between the uniform cumulative distribution and the observed one. Index (13), takes values in the range:

$$\frac{1-K}{2} \leq IS^3 \leq \frac{K-1}{2}.$$

and has the advantage of assuming negative values if the observed distribution presents worse judgements compared to the uniform distribution, assumes a value of zero if the two distributions are similar and assumes positive values if judgements are better than the totally indifferent distribution. Normalizing index IS^3, it is also trivial to verify that:

$$IS^3_{NORM} = \frac{IS^3 - \left(\frac{1-K}{2}\right)}{\frac{K-1}{2} - \frac{1-K}{2}} = 1 - \frac{1}{K-1}\sum_{k=1}^{K-1} F_k = IS^1 = IS^2 = IS. \quad (14)$$

This result highlights that even if index IS^3_{NORM} has a different genesis, nevertheless the behaviour of IS^1, IS^2 and IS^3_{NORM} (IS in the sequel) is exactly the same.

4 Weighting and aggregation

As highlighted in Section 2, after reducing the original data to numeric information through scoring transformations or by indices like those of Sections 3.1 and 3.2, the next two steps are weighting and aggregating such information. The *aggregating function* Ψ (also called *combining function* or *link function*) allows reducing data dimensionality mapping R^{2J} to R^1, *i.e.*: $\Psi : R^{2J} \to R^1$ where arguments of Ψ are the $q_j (j = 1, \ldots, J)$ values obtained by transforming elementary indicators and the $w_j (j = 1, \ldots, J)$ set of the J weights associated with each indicator.

The aggregating function assumes the form $CI = \Psi(q_1, \ldots, q_J; w_1, \ldots, w_J)$. As pointed out by several authors, the aggregating function should satisfy at least three minimal proprieties (see Lago and Pesarin, 2000; Arboretti Giancristofaro *et al.*, 2007), which define a class of functions. Some functions that satisfy proprieties are additive combining function, Fischer's combining function, Liptak combining function, logistic combining function and Tippet combining function. All these combining functions are nonparametric with respect to the underlying dependence structure among elementary indicators, since "*no explicit model for this dependence structure is needed and no dependence coefficient has to be estimated directly from data*" (Lago and Pesarin, 2000, p. 45). Folks (1984) and Pesarin (1992, 1994) discuss some characteristics of such combining functions, showing that they vary in their capacity to discriminate different situations to the same extent throughout the range of their values. One of the most frequently used aggregating function is the additive combining function (Saisana *et al.*, 2005), also called the *simple linear weighted function* or *simple additive weighting.* Besides its simplicity, this procedure is based on the Classical Test Theory that if the variables are parallel measurements (*i.e.*, are homogeneous), their sum will tend to cancel out measurement errors (Carpita and Manisera, 2012). Other authors suggest adopting, as the aggregate function, one element of the family of the generalized weighted mean of order s (Zani *et al.*, 2012). Notice that, in this case, if $s = 1$, the aggregate function is the additive combining function. Another method suggested is the *weighted product method* that corresponds to the weighted geometric mean. Some authors (Ebert and Welsch, 2004) show the advantages of this noncompensatory combining function with the assumption that only the ordinal characteristic of the CI is concerned, while Zhou and Zhou (2010) extend the weighted product methods and propose a multiplicative optimization approach that requires no prior knowledge of the system of weights.

Since all combining functions also have a vector of weights as arguments, one of the more debated issues regards how to choose these weights. This is difficult since no objective and unique criteria have emerged to solve the problem. Some criteria are derived from very complex multivariate methods, others from simpler but more subjective procedures, so that the different weighting schemes are arbitrary or unreliable (Cox *et al.*, 1992). Some general weighting criteria are:

- *equal weighting*, which imposes that all elementary indicators have the same weight. If equal weights are chosen, it is assumed that selected indicators balance different aspects of the unobservable variable;
- *weights obtained by expert judgement*, in this case, the weighting system is delegated to a nonstatistical sphere and can be obtained by the simple attributions by one or more persons considered expert regarding the object of the analysis. A more sophisticated method can also be used, such as the *participatory approach*, like *budget allocation* (Moldan *et al.*, 1997), or *analytic hierarchical processes* (Saaty, 1980; Forman, 1983) or some iterative procedure that starts from initial weights assigned by experts to try to maximize (or minimize) some suitable functions directly (or indirectly) linked to a variable's importance;
- *factor loadings*, obtained by principal component analysis (PCA) for quantitative variables or by nonlinear PCA (NL-PCA) for ordinal variables. In these cases, weights are proportional to the correlation between each elementary indicator and the component considered as a proxy of the latent variable (Carpita and Manisera, 2011; Ferrari and Barbiero, 2012; Zani and Berzieri, 2008);
- *normalized regression coefficients*, in this case, a single output indicator is selected and a multiple regression model is constructed to calculate the relative weights of the others indicators (Saisana and Tarantola, 2002);
- *weighting through fuzzy membership function*, which enables associating to each variable a weight that is an inverse function of the fuzzy proportion of each variable (Cerioli and Zani, 1990).

In this paper, we suggest two other methods to obtain a system of weights:

1. weighting through the Kendall association index;
2. weighting through the heterogeneity index.

4.1 Weighting through the Kendall association index

Like PCA and NL-PCA, in this case weights are chosen as a function of the level of association between each indicator and the unobservable variable. To obtain such association measures, there are two possibilities: (a) a proxy of the latent variable is available, then weights are relative measures of association of each indicator with this observed proxy variable. There are many applicative situations in which this proxy variable is available (for example, in customer/user satisfaction measures, generally there is an overall item that can be used for this); (b) if the proxy variable is not available, the weights are selected by considering the mean association between each observed indicator and all the others.

Given the ordinal level of indicators considered in this paper, the Kendall (1938) index τ has been selected to measure the association between elementary indicators. Indicating with $\tau_{j,p}$ the association between indicator q_j and proxy indicator q_p, the weight in case (a) will be:

$$w_j = \frac{\tau_{j,p}}{\sum_{r=1}^{J} \tau_{r,p}}, \ \tau_{j,p} = \frac{N_c^{j,p} - N_d^{j,p}}{\sqrt{(N_c^{j,p} - N_d^{j,p} + T_j)(N_c^{j,p} - N_d^{j,p} + T_p)}} \tag{15}$$

where: $N_c^{j,p}$ is the number of concordant pairs; $N_d^{j,p}$ is the number of discordant pairs; T_j is the number of pairs tied only on the q_j elementary indicator; T_p is the number of pairs tied only on the q_p proxy indicator.

If no proxy indicator is available (case b), the weights will be:

$$w_j = \frac{t_j}{\sum_{r=1}^{J} t_r}, \ t_j = \frac{1}{J-1} \sum_{h=1}^{J} \tau_{j,h} \quad \forall h = 1, \ldots, J \wedge h \neq j \tag{16}$$

4.2 Weighting through the heterogeneity index

The second method we suggest reflects the idea that each elementary indicator is more reliable if judges give more concordant opinions. On the contrary, if the distribution of the responses is very heterogeneous, it seems reasonable to give less relevance to that indicator because of its low reliability. These points lead us to consider a reverse function of the heterogeneity measure of each indicator as a system of weights. Taking as a measure of heterogeneity, for item j, the following index of dispersion for an ordinal variable (Leti, 1983):

$$d_j^* = \frac{4}{K-1} \sum_{k=1}^{K-1} F_{jk}(1 - F_{jk}),$$

it is possible to associate to each elementary indicator the weight:

$$w_j = 1 - \frac{d_j^*}{\sum_{r=1}^{J} d_r^*}. \tag{17}$$

In the next section, an application using a weighting system obtained with (15), (16) and (17) is considered and compared.

5 An application to teaching performance of university courses

As an illustration of the proposed methods defined in Sections 3 and 4, we will build the ‘teaching performance index’, a CI to evaluate the teaching service of a set of university courses. The data refer to questionnaires of course evaluations administrated to the students of the Faculty of Economics, University of Udine, during the first semester of the academic year 2010–2011.

The total number of questionnaires included in this analysis is 365, corresponding to seven compulsory courses. The number of questionnaires without missing data ranged from a minimum of 30 to a maximum of 60. For privacy reasons, the names of the courses are substituted with labels. The standard questionnaire has 19 items divided into three sections: teaching performance (12 items), organizational aspects (6 items) and global satisfaction (one item). All the items have the same ordinal responses: 1=not at all satisfied, 2=dissatisfied, 3=satisfied, 4=very satisfied. In this framework, students (judges) express their opinion about different aspects (items) of the teaching service. The aim is to obtain, starting from simple item indicators, a CI to evaluate teaching performance among courses.

As stated in Section 2, there are five steps in constructing a CI. In this analysis we consider the following solutions to select (step 1), transform (step 2), weight (step 3) and aggregate (step 4) the elementary item indicators.

Step 1. As the aim is to obtain a “teaching performance index”, we select 12 elementary indicators, *i.e.* the items related to teaching performance.

Step 2. Responses, coded from 1 to 4, are scored with three different methods, so three different elementary item indicators are calculated.

Method A. The dissimilarity index IS of formula (14) ;

Method B. Symmetric scoring system (see formula (1)), separating satisfaction judgements (3=satisfied, 4=very satisfied) from dissatisfaction judgements (1=not at all satisfied, 2=dissatisfied). For item j, the elementary item indicator is obtained with formula (2);

Method C. Asymmetric scoring system (see formula (3)), similar to the previous method separating satisfaction judgements from dissatisfaction judgements. For item j, the elementary item indicator is obtained with formula (4).

In order to compare the elementary indicators among items, we consider further transformations. As the dissimilarity index IS is normalized, its values are directly comparable. With the purpose of transforming the values of elementary indicators based on a symmetric and asymmetric scoring system, we use the definition of “strong performance profile” described in Section 3.1 to find the minimum and maximum corresponding to the lowest level and the highest level, respectively, of teaching performance.

To transform indicators, we consider the transformation functions described in Section 3.1. As we have four levels of satisfaction, we code these with $v_1 = 1, v_2 = 2, v_3 = 3$ and $v_4 = 4$; after transformations (7), (8) and (9), we obtain the new indicators summarized in Table 1.

Table 1: Elementary item indicators and different types of score transformation

Scoring Method	Minimum	Maximum	Transformation		
			a	b	c
Symmetric	0.5	4	$\bar{s}_j^{*S} = 1 - \frac{0.5}{\bar{s}_j^S}$	$\bar{s}_j^{*S} = \frac{\bar{s}_j^S}{4}$	$\bar{s}_j^{*S} = \frac{\bar{s}_j^S - 0.5}{3.5}$
Asymmetric	1	4.5	$\bar{s}_j^{*A} = 1 - \frac{1}{\bar{s}_j^A}$	$\bar{s}_j^{*A} = \frac{\bar{s}_j^A}{4.5}$	$\bar{s}_j^{*A} = \frac{\bar{s}_j^A - 1}{3.5}$

There are seven transformed item indicators: one with the dissimilarity index, three with the symmetrical scoring system and three with the asymmetrical scoring system.
Step 3. The systems of weights are defined with (15), (16) and (17) and denoted, respectively, with $w_{j,1}$, $w_{j,2}$ and $w_{j,3}$, $j = 1, 2, \ldots, 12$.
Step 4. The global teaching performance index TPI_i for the $i-th$ course is built using the additive combining function:

$$TPI_i = \sum_{j=1}^{12} w_{j,v} \cdot q_j,\ \ v = 1, 2, 3;\ \ i = 1, 2, \ldots, 7$$

where q_j is one of the transformed elementary indicators for item $j - th$, illustrated in Step 2.

As we define seven types of transformed elementary indicators, three systems of weights and one combining function, for each course we have 21 global teaching performance indices. Applying the three steps illustrated above to the data for each course, the values of the TPI indices allow us to rank courses on the basis of teaching performance. Results summarized in Table 2 show the ranking of courses (rank 1 corresponds to the best course and 7 to the worst course).

In particular, table 2 shows that differences in ranking position, due to different indices with a different weighting system, vary from 1 (Course 2 and Course 4) to 4 (Course 1). Notice that Course 2 and Course 4 are almost always in the highest ranking while Course 3 is generally in the lowest.

Even if the application is illustrated for a limited number of courses, it is important to notice that this procedure is useful in different ways because it is simple, allows to detect clusters of courses that share a similar level of student satisfaction and to disseminate the results in a useful way.

6 Concluding remarks

The use of CI is useful when the focus is mainly on a comparison of performances. This leads typically to a ranking of the evaluated units. In this paper, we presented different approaches to transform ordinal data into quantitative data, to construct elementary indicators and to summarize them in a unique CI useful for evaluating the performance

Table 2: Teaching performance - Course ranking

Index	Course 1	Course 2	Course 3	Course 4	Course 5	Course 6	Course 7
TPI1	3	1	7	2	6	4	5
TPI2	5	1	7	2	6	3	4
TPI3	5	1	7	2	6	3	4
TPI4	7	1	6	2	5	3	4
TPI5	7	1	6	2	5	3	4
TPI6	7	1	6	2	5	3	4
TPI7	3	1	7	2	6	4	5
TPI8	5	1	7	2	6	3	4
TPI9	5	1	7	2	6	3	4
TPI10	7	1	6	2	5	3	4
TPI11	7	1	6	2	5	3	4
TPI12	7	1	6	2	5	3	4
TPI13	5	1	7	2	6	4	3
TPI14	5	1	7	2	6	4	3
TPI15	7	1	6	2	5	4	3
TPI16	5	1	7	2	6	4	3
TPI17	3	1	7	2	6	5	4
TPI18	5	1	7	2	6	4	3
TPI19	5	1	7	2	6	4	3
TPI20	5	2	7	1	6	4	3
TPI21	5	2	7	1	6	4	3

$TPI1 - TPI3$, indices based on dissimilarity index IS^1; $TPI4 - TPI12$, indices based on symmetrical scoring; $TPI13 - TPI21$, indices based on asymmetrical scoring.

of evaluated units. Particular emphasis is given to the use of the cumulative function to obtain elementary indicators and to different weighting systems, based on dispersion indices and on Kendall association index. Finally, the proposed procedures are applied to data about teaching performance. Differences in ranking position are very small for courses that are in the highest and in the lowest part of the rank, while the differences are greater for courses in the middle.

References

Arboretti Giancristofaro, R., Bonnini, S. and Salmaso, L. (2007) A performance indicator for multivariate data. *Quaderni di Statistica*, LI:1–29.

Bartholomew, D.J. (1987). *Latent variables models and factor analysis.* Griffin & Co. Limited, London.

Bollen, K.A. (1989). *Structural equations with latent variables.* Wiley & Sons, New York.

Borsboom, D., Mellenbergh, G.J. and Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110(2):203–219.

Capursi, V. and Porcu, M. (2001). La didattica universitaria valutata dagli studenti: un indicatore basato su misure di distanza fra distribuzioni dei giudizi. In *Atti Convegno Intermedio SIS "Processi e metodi statistici di valutazione"*, Roma, 4-6 giugno 2001.

Carpita, M. and Manisera, M. (2011). On the nonlinearity of homogeneous ordinal variables. In S. Ingrassia, R. Rocci (eds.) *New perspectives in statistical modeling and data analysis*. Springer, Heidelberg, pages 489–496.

Carpita, M. and Manisera, M. (2012). Constructing indicators of unobservable variables from parallel measurements. *Electronic Journal of Applied Statistical Analysis*, 5(3):320–326.

Cerioli A. and Zani, S. (1990). A fuzzy approach to the measurement of poverty. In C. Dagum and M. Zenga (eds.), *Income and wealth distribution, inequality and poverty*, studies in contemporary economics. Springer Verlag, Berlin, pages 272–284.

Cox, D., Fitzpatrik, R., Fletcher, A., Gore, S., Spiegelhalter, D. and Jones, D. (1992). Quality-of-life assessment: can we keep it simple? *Journal of Royal Statistical Society*, 155(3):353–393.

D'Elia, A. and Piccolo, D. (2005). A mixture model for preference data analysis. *Computational Statistics & Data Analysis*, 49:917–934.

Ebert, U. and Welsch, H. (2004). Meaningful environmental indices: a social choice approach. *Journal of Environmental Economics and Management*, 47:270–283.

Ferrari, P.A. and Barbiero, A. (2012). Nonlinear principal component analysis. In: R.S. Kennet and S. Salini (eds): *Modern analysis of customer survey: with applications using R*, J. Wiley & Sons, Ltd, Chichester, pages 333–356.

Folks, J.L. (1984). Combinations of independent tests. In P.R. Krishnaiah and P.K. Sen (eds.) *Handbook of statistics, 4*. Elsevier Science Publisher, Amsterdam.

Forman, E.H. (1983). The analytic hierarchy process as a decision support system. In *Proceedings of the IEEE Computer Society*.

Gifi, A. (1990). *Nonlinear multivariate analysis*. Wiley & Sons, Ltd, Chichester.

Iannario, M. and Piccolo, D.(2010). A new statistical model for the analysis of customer satisfaction. *Quality Technology & Quantitative Management*, 7(2):149–168.

Joreskog, K. and Moustaki, I. (2001). Factor analysis for ordinal variables: a comparison of three approaches. *Multivariate Behavioural Research*, 36:347–387.

Kendall, M. (1938). A new measure of rank correlation. *Biometrika*, 30:81-89.

Kenett, R. and Salini, S., editors (2012). *Modern analysis of customer surveys: with applications using R*. John Wiley & Sons, Ltd, Chichester.

Lago, A. and Pesarin, F. (2000). Nonparametric combination of dependent rankings with application to the quality assessment of industrial products. *Metron*, LVIII:39–52.

Lauro, C. and Nappo, D. (2011). Model based composite composite indicators. In ISTAT Workshop on *La misurazione di fenomeni multidimensionali: indici sintetici ed esperienze a confronto*, Rome, March 2, 2011. Available online at: `ttp://www3.istat.it/istat/eventi/2011/worksop2marzo/`Lauro.pdf.

Leti, G. (1983). *Statistica descrittiva*. Il Mulino, Bologna.

Moldan, B., Billharz, S. and Matravers, R. (1997). *Sustainability indicators: report of the projection on indicators of sustainable development*. Wiley & Sons, Ltd, Chichester.

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its application*. University of Toronto Press, Toronto.

OCSE (2008). *Glossary of Statistical Terms*. Available online at: `http://stats.oecd.org/glossary`

Pesarin, F. (1992). A resampling procedure for nonparametric combination of several dependent tests. *Journal of the Italian Statistical Society*, 1:87–101.

Pesarin, F. (1994). Goodness of fit testing for ordered discrete distributions by resampling techniques. *Metron*, LII:57–71.

Piccolo, D. (2009). A class of models for ordinal variables: logical foundations and statistical issues. In *Proceedings of the conference "Multivariate Methods and Models for evaluating Public Services"*, Book of Abstract, Rimini.

Portoso, G. (2003). L'esponenziale e la normale nella quantificazione determinata indiretta: un indicatore d'uso. *Rivista Italiana di Economia Demografia e Statistica*, LVII:135–139.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research, Copenhagen.

Saaty, T.L. (1980). *The analytic hierarchy process*. McGraw-Hill, New York.

Saisana, M., Saltelli, A. and Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society, Series A*, 168(2):307–323.

Saisana, M. and Tarantola, S. (2002). State-of-the-art report on current methodologies and practices for composite indicator development, EUR 20408 EN, European Commission-JRC, Italy.

Skrondal, A. and Rabe-Hesketh, S. (2007). Latent variable modelling: a survey. *Scandinavian Journal of Statistics*, 34:712–745.

Thurstone, L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33:529–554.

Torgerson, W.S. (1958). *Theory and methods of scaling*. Wiley, New York.

Zanarotti, M.C. (2012). Scelta della distribuzione di riferimento nelluso degli indici di dissomiglianza per la valutazione con dati ordinali. *Serie EPN 141*,n.11, Università Cattolica Sacro Cuore, Milano.

Zani, S. and Berzieri, L. (2008). Measuring customer satisfaction using ordinal variables: an application in a survey on a contact center. *Statistica applicata*, 20:331–351.

Zani, S., Milioli, M.A., and Morlini, I. (2012). Fuzzy methods and satisfaction indices. In R.S. Kennet and S. Salini (eds): *Modern analysis of customer survey: with applications using R*, Wiley & Sons, Ltd, Chichester.

Zhou, P., Ang, B.W. and Zhou, D. Q. (2010). Weighting and aggregation in composite indicator construction: a multiplicative optimization approach. *Social Indicator Research*, 96:169–181.